



Draft Genome Sequence of *Corynebacterium sanguinis* Strain Marseille-P8776

Mudra Khare,^{a,b} Dhiraj Sinha,^{a,b} Rita Zgheib,^{a,b} Amael Fadlane,^{a,b}  Didier Raoult,^{b,c}  Pierre-Edouard Fournier^{b,c}

^aAix Marseille University, Institut de Recherche pour le Développement (IRD), Service de Santé des Armées, Assistance Publique-Hôpitaux de Marseille, UMR Vecteurs Infections Tropicales et Méditerranéennes (VITROME), Institut Hospitalo-Universitaire Méditerranée Infection, Marseille, France

^bInstitut Méditerranée-Infection, Marseille, France

^cAix-Marseille Université, Institut de Recherche pour le Développement (IRD), UMR Microbes Evolution Phylogeny and Infections (MEPHI), Institut Hospitalo-Universitaire Méditerranée-Infection, Marseille, France

ABSTRACT In 2018, *Corynebacterium sanguinis* strain Marseille-P8776 was isolated from the blood of a 64-year-old woman suffering from breast cancer who had undergone chemotherapy and radiotherapy. Following whole-genome sequencing, the chromosome of strain Marseille-P8776 was 2,613,836 bp long with a G+C content of 64.92%, 2,568 protein-coding genes, and 64 RNA genes.

The genus *Corynebacterium* was created by Lehmann and Neumann in 1896, but the main features of this genus were first described by Collins and Cummins in 1986 (1). Members of the genus are distributed in various places such as soil, water, plants, and food products, and some species are commensals of mucosa and skin from humans and animals (2). Currently, the genus *Corynebacterium* is composed of 136 species (<https://lpsn.dsmz.de/genus/corynebacterium>) (3). In 2018, we isolated strain Marseille-P8776 from a blood sample taken from a 64-year-old woman suffering from breast cancer and being treated with radiotherapy and chemotherapy at La Timone Hospital (Marseille, France).

Bacterial growth was obtained following culture on 5% sheep blood-enriched Columbia agar (bioMérieux, Marcy l'Etoile, France) in an aerobic atmosphere at 37°C for 24 h. Genomic DNA (gDNA) from strain Marseille-P8776 was quantified using a Qubit assay (Life Technologies, Carlsbad, CA, USA) to 0.2 ng/μL. Next, the gDNA was sequenced using a MiSeq sequencer with the paired-end strategy and barcoded using the Nextera XT DNA sample prep kit (Illumina Inc., San Diego, CA, USA).

The genomic DNA was diluted to obtain 1 ng as input. In the tagmentation step, the DNA was fragmented and tagged. Then, using limited-cycle PCR amplification (12 cycles), the tag adapters were completed, and dual-index barcodes were introduced. After purification on AMPure XP beads (Beckman Coulter Inc., Fullerton, CA, USA), the libraries were normalized on specific beads according to the Nextera XT protocol (Illumina). The pooled single-strand library was loaded onto the reagent cartridge and then onto the instrument along with the flow cell. Automated cluster generation and paired-end sequencing with dual index reads were performed in a single 39-h run in 2 × 250-bp format.

Total information of 3.9 Gb was obtained from a 491 K/mm² cluster density with a cluster passing quality control filters of 93.50%. Within this run, the index representation of strain Marseille-P8776 was 3.83%. The 9,478,951 paired-end reads were filtered according to the read quality using FastQC version 0.11.8 with default parameters (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). The genome assembly was performed as previously described (4) using the SPAdes version 3.11.1 software (5) that includes GapCloser and PEResolver for gap and overlap removal (6). The “careful” option was used to minimize the number of mismatches and short indels. Default parameters were applied for k values, i.e., k-mer values of 21, 33, 55, 77, 99, and 127. The

Editor Julie C. Dunning Hotopp, University of Maryland School of Medicine

Copyright © 2022 Khare et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Pierre-Edouard Fournier, pierre-edouard.fournier@univ-amu.fr.

The authors declare no conflict of interest.

Received 25 January 2022

Accepted 11 April 2022

Published 4 May 2022

genome sequence was annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) version 4.12 (7). The predicted protein sequences were compared to the Clusters of Orthologous Groups (COGs) database using BLASTP (E value, $1e-03$; coverage, 0.7; and identity, 30%). Strain Marseille-P8776 had a genome size of 2,613,836 bp assembled into one scaffold with a G+C content of 64.92%. Of 2,847 predicted genes, 2,691 were protein-coding genes and 63 RNAs (4 rRNAs, 55 tRNAs, and 4 noncoding RNAs [ncRNAs]). A total of 1,937 genes (73.59%) were assigned a putative function by COGs, and 695 genes (26.40%) were annotated as hypothetical proteins.

Using ABRicate software (8), the genomic sequences were screened against the CARD (Comprehensive Antibiotic Resistance Database) (9) for resistance genes and the VFDB (Virulence Factors Database) (10) for virulence factors, revealing that strain Marseille-P8776 harbored 6 resistance genes and 7 virulence genes.

Data availability. The draft genome sequence of *C. sanguinis* strain Marseille-P8776 has been deposited at GenBank under accession number [NZ_JACEOR000000000.1](https://ncbi.nlm.nih.gov/GenBank/record/NZ_JACEOR000000000.1) and the raw reads under SRA accession number [SRR17477869](https://ncbi.nlm.nih.gov/SRA/record/SRR17477869) (BioProject accession number [PRJNA645897](https://ncbi.nlm.nih.gov/BioProject/record/PRJNA645897) and BioSample accession number [SAMN15518506](https://ncbi.nlm.nih.gov/BioSample/record/SAMN15518506)).

ACKNOWLEDGMENTS

The study was supported by the Méditerranée Infection Foundation, National Research Agency, under the program “Investissements d’avenir” (reference ANR-10-IAHU-03), and by Région Provence Alpes Côte d’Azur and European funding FEDER PRIM1.

We declare no competing interests.

M.K. performed the genomic analysis and drafted the article; D.S. participated in the genomic analysis; R.Z. participated in drafting the manuscript; A.F. isolated and characterized the strain; D.R. and P.-E.F. designed the study and drafted the article. All authors reviewed and validated the final manuscript.

REFERENCES

- Collins MD, Cummins CS. 1986. Genus *Corynebacterium*, p 1266–1276. In Sneath PHA, Mair NS, Sharpe HE, Holt JG (ed), *Bergey's manual of systematic bacteriology*, vol 2. Williams & Wilkins, Baltimore, MD.
- Yassin AF, Kroppenstedt RM, Ludwig W. 2003. *Corynebacterium glaucum* sp. nov. *Int J Syst Evol Microbiol* 53:705–709. <https://doi.org/10.1099/ijs.0.02394-0>.
- Woese CR. 1987. Bacterial evolution. *Microbiol Rev* 51:221–271. <https://doi.org/10.1128/MMBR.51.2.221-271.1987>.
- Abou Abdallah R, Cimmino T, Baron S, Cadoret F, Michelle C, Raoult D, Fournier P-E, Bittar F. 2017. Description of *Chryseobacterium timonianum* sp. nov., isolated from a patient with pneumonia. *Antonie Van Leeuwenhoek* 110:1121–1132. <https://doi.org/10.1007/s10482-017-0885-8>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Xu M, Guo L, Gu S, Wang O, Zhang R, Peters BA, Fan G, Liu X, Xu X, Deng L, Zhang Y. 2020. TGS-GapCloser: a fast and accurate gap closer for large genomes with low coverage of error-prone long reads. *Gigascience* 9: giaa094. <https://doi.org/10.1093/gigascience/giaa094>.
- Tatusova T, DiCuccio M, Badretdin A, Chetvermin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>.
- Arnott A, Wang Q, Bachmann N, Sadsad R, Biswas C, Sotomayor C, Howard P, Rockett R, Wiklendt A, Iredell JR, Sintchenko V. 2018. Multi-drug-resistant *Salmonella enterica* 4,[5],12:i:- sequence type 34, New South Wales, Australia, 2016–2017. *Emerg Infect Dis* 24:751–753. <https://doi.org/10.3201/eid2404.171619>.
- Jia B, Raphenya AR, Alcock B, Waglechner N, Guo P, Tsang KK, Lago BA, Dave BM, Pereira S, Sharma AN, Doshi S, Courtot M, Lo R, Williams LE, Frye JG, Elsayegh T, Sardar D, Westman EL, Pawlowski AC, Johnson TA, Brinkman FSL, Wright GD, McArthur AG. 2017. CARD 2017: expansion and model-centric curation of the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res* 45:D566–D573. <https://doi.org/10.1093/nar/gkw1004>.
- Chen L, Zheng D, Liu B, Yang J, Jin Q. 2016. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res* 44: D694–D697. <https://doi.org/10.1093/nar/gkv1239>.